



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Systems biology, synthetic biology and data-driven research:

**Citation for published version:**

Calvert, J 2012, 'Systems biology, synthetic biology and data-driven research: a commentary on Krohs, Callebaut, and O'Malley and Soyer', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, no. 1, pp. 81-84.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences

**Publisher Rights Statement:**

© Calvert, J. (2012). Systems biology, synthetic biology and data-driven research:: a commentary on Krohs, Callebaut, and O'Malley and Soyer. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1), 81-84.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## **Systems biology, synthetic biology and data-driven research: a commentary on Krohs, Callebaut, and O'Malley and Soyer**

Jane Calvert

Published as: Calvert, J (2012) 'Systems biology, synthetic biology and data-driven research: a commentary on Krohs, Callebaut, and O'Malley and Soyer' *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 81–84.

The three papers that I will discuss here all focus on systems biology. This is significant in the context of data-driven research, because systems biology is the approach to biology which explicitly attempts to make sense of the vast amounts of 'omics' data generated by high-throughput techniques. Two of the three papers (Callebaut and O'Malley and Soyer) additionally address synthetic biology. Systems and synthetic biology have much in common, and both can be interpreted as attempts to deliver on the promises of the genome sequencing projects, in respect to both biological understanding and applications (Calvert and Fujimura, 2011). The two fields do have important differences in their orientation and aims, however, and I will return to this point in my discussion of engineering below.

The three papers came out of a workshop on 'Data-driven research in the biological and biomedical sciences'. The workshop was motivated by questions such as: Is data-driven research leading to changes in the ways in which biology is done? How should we understand these changes? What epistemological issues do they raise? Where do they leave hypothesis-driven research? And how do these developments challenge current thinking in the philosophy, history and sociology of science? As I understood it, the overarching aim of the workshop was to attempt to "come to grips philosophically with the transformations of biology in this century" (Callebaut p.14).

My aim in this discussion paper is to start a conversation between the three papers. And something immediately notable about all of them is that they introduce new (or relatively new) concepts, such as convenience experimentation (Krohs), exploratory questioning (O'Malley and Soyer), scientific perspectivism (Callebaut), and integration (O'Malley and Soyer). This introduction of new concepts suggests that we are seeing changes in the way in which biology is done that we currently do not have the conceptual tools to grasp. These papers seek to provide us with new tools to think with, to give us a better understanding of emerging biological practices. All the papers also contain (either implicitly or explicitly) the view that categorising biological research as either hypothesis-driven or data-driven is not satisfactory, and that explaining current developments in terms of a shift toward data-driven research somehow does not capture everything we want to explain.

I start by discussing Krohs' argument about the over-reliance by top-down systems biology on convenience experimentation. This paper demonstrates the importance of technological developments and how they influence what comes to be known. I then compare Krohs' discussion of exploratory experimentation with O'Malley and Soyer's notion of exploratory questioning, and I go on to look at the other epistemic

features of systems biology highlighted by O'Malley and Soyer, focusing particularly on integration, which I think is a crucially important concept. I connect their discussion of integration to some of my own empirical work on interdisciplinarity in systems biology. I argue that Callebaut's scientific perspectivism has similarities to the integration discussed by O'Malley and Soyer. I then reflect on one strand of Callebaut's article, dealing with synthetic biology, to put forward some ideas about the importance of taking engineering seriously in the social and philosophical studies of the life sciences.

### **Krohs on convenience experimentation**

Krohs' paper argues that there is a strong reliance on what he calls convenience experimentation in top-down systems biology, a branch of systems biology that studies 'omic' interactions at the whole cell scale. (As Krohs himself points out, top-down systems biology is only one strand of systems biology). Krohs summarizes his argument: "in convenience experimentation many experiments are done in the way they are actually done for the reason that they are so extraordinary convenient to perform" (p.13). He says that convenience experimentation has developed in a context where data has become extremely plentiful, and argues it is convenience experimentation, rather than hypotheses, that drives model building in top-down systems biology. This has important implications, because practicing science in such a manner "strongly channels research" (p.3). To use a familiar analogy, like a drunk who looks for his lost keys under the lamp-post, because that is where the light is, Krohs' argument is that scientists do certain experiments simply because they have the technologies available (such as microarrays, for example). This type of experimentation carries low epistemic risk, as illustrated by a different analogy Krohs uses in a footnote: "The situation resembles industrial prefabrication of meals, convenience food, which simplifies home cooking – and standardizes its outcome" (p.3).

Krohs elucidates the nature of convenience experimentation by drawing a comparison between metabolic pathway analysis and top-down systems biology. In an earlier paper he explains that top-down systems biology is concerned with the topology of a network "without further characterizing the components in any other way than by describing their place within the network (as a "node" of the network) and the interactions they are engaged in ("edges" in the terminology of network analysis)" (Krohs 2010, p. 154-155). He explains that while metabolic pathway analysis assumes a 'general explanatory hypothesis of localisation', where each reaction has an identifiable role in a metabolic pathway, top-down systems biology adopts a 'general explanatory hypothesis of delocalization' where enzymes do not have stable roles, but are simply regarded as nodes in a broader network, meaning that regulatory functions are assumed to be delocalized. He argues that the methodology of convenience experimentation in top-down systems biology will only produce datasets that satisfy "the preconception of delocalized functionality" (p.11). The tools and models adopted strongly influence the conclusions that can be drawn.

He also maintains that the kind of datasets used in top-down systems biology have led to a change in modelling strategies towards Boolean networks, "and

consequently to a change in the epistemic goals that can be followed” (Krohs p.8). This profound epistemic shift in the way that science is done is what is particularly interesting about Krohs’ analysis of convenience experimentation. He says we are seeing dramatic changes in biological theorizing which “follow in a clandestine manner” (Krohs p.3) from convenience experimentation.

The situation is different with data-mining, however, which, it could be argued, is an important part of systems biology. Krohs says “when data mining is also taken into consideration, convenience experimentation based research becomes exploratory” (p.11). The word ‘exploratory’ is important here, because Krohs distinguishes between three types of experimentation: convenience experimentation, hypothesis-driven experimentation, and exploratory experimentation (drawing on Burian 1997 and Steinle 1997). He describes exploratory experimentation as the kind of research that requires “uncertainty about the conceptual framing of the experiment, its relevance, or even about the very phenomenon to be investigated” (p.13). Experimentation is exploratory only when the researcher does not know what they are going to find out. If Krohs’ arguments hold, this is clearly not the case with convenience experimentation.

### **O’Malley and Soyer on integration**

I think a helpful way of understanding Krohs notion of exploratory experimentation is by linking it to O’Malley and Soyer’s discussion of ‘exploratory questions’. They describe these as being broad general questions rather than specific hypotheses. O’Malley and Soyer give some examples of these, including ‘what if’ questions (p.17), and questions like ‘what’s going on here, and what happens when we construct things differently?’ (p.21).

According to O’Malley and Soyer, exploratory questioning is only one element of integration, and integration is the most important analytical category in their paper. They divide integration into three components: exploratory questioning, technological development and the transfer of explanations from one research domain to another. They argue convincingly that these factors are of more importance in understanding scientific change than an emphasis on data-driven or hypothesis-driven approaches. In fact, they suggest that we should break away from the classification of research as either hypothesis- or data-driven towards understanding science in terms of a more inclusive range of practices. This perspective is refreshing, and also demonstrates the need to rethink our existing conceptual categories in the light of developments in the life sciences.

They describe integration as combining different datasets, methods, and approaches. They describe disciplinary integration as one of the conditions for integration, but I think we could see it as a form of integration itself, since systems biology requires that physicists, computer scientists, engineers, mathematicians, statisticians and biologists come together in new interdisciplinary configurations. The advantage of using ‘integration’ in this broader sense is that it works at many different levels: theoretical, methodological and social.

If we think of disciplinary integration as a form of integration, empirical work becomes relevant. O'Malley and Soyer note that for many biologists "optimal individuals are those who have trained in one discipline and learned to work with other disciplines" adding "Monodisciplinary training is still considered to be the best way to avoid the production of undertrained but multidisciplinary researchers" (p.28). My qualitative empirical study of 35 systems biologists shows there is not a consensus amongst people working in systems biology that monodisciplinary training is preferred, however, but there are a range of views about how future systems biologists should be educated (Calvert 2010). For example, some systems biologists think radical changes in science education are necessary, reaching down to undergraduate or even high school level, while other senior systems biologists argue that in the future all scientists should have a dual major. In one doctoral training centre for systems biology in the UK graduate students are trained to develop a specific area of expertise, but also to have the ability to talk across disciplines. In this way they develop a 'light' expertise in a discipline outside their own. Others hope the interdisciplinary training of undergraduates and postgraduates means that we will see a movement from specialists to 'integrators' in the future.

A distinction can also be made between collaborative interdisciplinarity (where individuals come together from different disciplines to work on a problem), and individual interdisciplinarity (where multiple disciplinary skills are found within one individual). Although, as O'Malley and Soyer point out, some commentators doubt the value of the latter, for others, training multi-skilled systems biologists is the long-term aim. This is currently more of an aspiration than a reality at the moment, although in some cases we are starting to see the 'wet' (lab) and 'dry' (computational) distinction breaking down, with some systems biologists starting to talk about being "moist" and even "soggy" and "damp" researchers, depending on the type of research they are doing (Calvert 2010).

It is not easy to introduce changes in the way scientists are trained, however, because these changes have the potential to destabilise existing disciplines and practices.<sup>1</sup> This destabilization may explain why systems biologists often face institutional resistance when setting up centres and institutes dedicated to systems biology. And once these institutes are set up, tensions will often arise because of competing ideas about what systems biology should achieve, and different attitudes towards quantification and predictability (see Calvert and Fujimura 2011).

These are small additions to O'Malley and Soyer's brief comments on interdisciplinarity in systems biology, and my aim in making these comments is to show that it may be useful to adopt a broader understanding of integration. In fact, such an understanding of integration has significance beyond the philosophy of systems biology. For example, the US's National Research Council's 2009 report *A New Biology for the 21st Century* argues "the essence of the New Biology is integration – re-integration of the many subdisciplines of biology, and the integration into biology of physicists, chemists, computer scientists, engineers, and

---

<sup>1</sup> This is an example of what Barry et al. (2008) call the 'agonistic-antagonistic' mode of interdisciplinarity.

mathematicians” (p.vii). Integration is starting to become an important analytical category not only in epistemic, methodological and social contexts, but also at the level of science policy.

To add a final further dimension, the interdisciplinary integration we see in fields like systems and synthetic biology may extend even further than the natural and physical sciences to incorporate the social sciences and humanities. Although some might assume that this would be a difficult stretch, the paper by O’Malley and Soyer is itself an excellent example of a collaboration between a philosopher and a systems biologist. This broader understanding of ‘integration’ connects to Fisher et al.’s (2006) use of the term to refer to the integration of perspectives from the social sciences and the humanities into science and engineering. It also ties into a point Callebaut makes in a footnote, that philosophers should not merely be running behind the scientists uncritically, but, as Francisco Varela puts it, “helping us by suggesting some wild ideas” (p.4).

### **Callebaut on scientific perspectivism**

The key concept in Callebaut’s paper on big-data biology is scientific perspectivism.<sup>2</sup> This is an idea I cannot do justice to here, but crudely put it is a philosophical position that recognises that we always perceive the world from a particular point of view because of factors such as our observational vantage point, our theoretical position, and even the language that we speak. Scientific perspectivism leads to the conclusion that reality itself is “multi-perspectival” (Callebaut p.18), and it is not possible to reduce these perspectives to a single meaning.<sup>3</sup> In other words, there is no ‘God’s eye view’.

Callebaut makes the interesting suggestion that complex systems are more likely than simple systems to “require the use of information from more than one perspective for their solution” (p.18). This implies that data-driven research, and systems biology in particular, will demand the integration of multiple perspectives. And systems biologists support this point when they argue that it is the complexity of the object (the biological system) that makes interdisciplinary collaboration a necessity in their field. The assumption that the more complex the object the broader the range of expertise required is what Mattila (2005) calls ‘object oriented interdisciplinarity’, where new objects of study lead to new interdisciplinary arrangements.

As will have probably already become apparent, Callebaut’s discussion of scientific perspectivism has very interesting resonances with O’Malley and Soyer’s discussion of integration. The interdisciplinarity which is a condition for both is highlighted by Callebaut when he says “biocomplexity research may require collaborations among disciplines as disparate as oceanography and epidemiology” (p.21).

---

<sup>2</sup> Callebaut draws on Giere (2006) here, but also Wimsatt (2007) and Van Fraassen (2008), among others.

<sup>3</sup> As a sociologist, I see links between scientific perspectivism and Haraway’s (1988) idea of situated knowledges.

Another feature Callebaut argues is important to scientific perspectivism is that it “should also fully take into account the collective, distributed nature of scientific cognition” (Callebaut p.22). This connects to the idea of collaborative interdisciplinarity discussed above. With collaborative interdisciplinarity one person does not need to have all the expertise necessary to deal with the diverse datasets and approaches being brought to bear in systems biology; instead this expertise is spread over interdisciplinary teams of researchers. We could argue that where multiple complex datasets are the focus we are likely to see a form of distributed cognition, where knowledge is shared amongst a heterogeneous scientific community.

### **Engineering life**

Scientific perspectivism is the central concept in Callebaut’s paper, but another important strand of the paper is exhibited in the idea of ‘engineering life’ as ‘changing the living world without trying to understand it’. This is a view that Callebaut draws from Woese’s (2004) famous article ‘A new biology for a new century’, where Woese laments about the current lack of a ‘guiding vision’ for the life sciences, saying that without such a vision science becomes an engineering discipline. This requires further analysis because neither Callebaut nor Woese explain what is meant by engineering. In fact, as Callebaut admits, he does not talk directly about what it means to engineer life, or about the aspiration to change the living world without trying to understand it.

Both Woese and Callebaut seem to be using ‘engineering’ as a placeholder for trends in the life sciences they are concerned about. However, most discussions of engineering do not describe it as being data-driven. Woese equates engineering with technological advance, but this characterisation overlooks distinctive features of engineering. Similarly, engineering may be mechanistic and reductionistic, as both Woese and Callebaut imply, but reductionism is not the motivation behind engineering.

I think that the nature of engineering is a very interesting topic that deserves further investigation by philosophers, sociologists and historians of science (particularly those who are interested in synthetic biology). Work has already been done on this topic of course, and one of the most well-known contributors is Vincenti (1990), who describes engineering as “the practice of organizing the design and construction of any artifact which transforms the physical world around us to meet some recognized need” (p.6). Rather than being focused on technology development or data accumulation, in engineering the emphasis is on meeting recognised needs. In this sense, knowledge is a means to a certain end (not an end in itself) for engineers. Vincenti even says “Engineering can, in fact, be defined in terms of these ends” (p.6).

This instrumentalism might itself explain why engineering seems to jar with thinkers like Woese and Callebaut. Woese characterises science as “an endless search for truth”, and I would agree that this is certainly not the aim of engineering, but this is because engineering is different from science. The instrumentalism of engineering, and the assumed superiority of ‘head’ over ‘hand’ might also explain why

engineering is a topic that has been neglected by the philosophy of science, which is particularly concerned with the acquisition of knowledge. The aim of engineering is not to increase our understanding of the world, but to change it (to paraphrase Marx). It is not a criticism of engineering to say that it is instrumental, because the whole point of engineering is to put scientific knowledge to practical uses.

I think a study of synthetic biology, in particular, benefits from taking engineering seriously; not condemning it for not furthering the pursuit of knowledge, but recognising that it has different aims. Keller (2009) acknowledges this, but argues that since the guiding aim of synthetic biology is not to find out about the natural world, it should not be called 'biology', a point that Woese and Callebaut might agree with.

If we understand synthetic biology as a branch of engineering, as some of its proponents maintain, then it becomes hard to put synthetic biology under the broad heading of data-driven research. Synthetic biology may have been one of the disciplines that emerged in the wake of the data generated by the genome projects, but the aim of the field is to make biological devices that perform desired functions. Synthetic biologists, unlike other post-genomic scientists, do not talk about drowning in data; they talk about not being able to get their constructions to work.

I recognise that not all of those who do research under the heading of 'synthetic biology' see their work as a branch of engineering, and in my depiction of synthetic biology here I am drawing primarily on the parts-based approaches (see for example Heinmann and Panke 2006; Endy 2005; Adrianantanandro et al. 2006; Brent 2004). I also recognise that much that goes on under the heading of 'synthetic biology' aims to increase our understanding of biological systems. Some fascinating examples of this type of synthetic biology are given by O'Malley and Soyer in their discussion of noise biology. But I do think the engineering approach to biology has novel conceptual consequences, which deserve further investigation in their own right. For example, one of the features of engineering that is much discussed in engineering-oriented branches of synthetic biology is the engineering design cycle (Royal Academy of Engineering 2009), and this has many interesting similarities to the iterativity discussed by O'Malley and Soyer.

### **Conclusions: new ways of thinking about scientific practice**

All three of the papers discussed here started from an interest in the transformations in biology that are associated with data-driven research, particularly those manifested by systems biology. Something which underlies all the papers is the importance of technological changes in driving scientific research. All the papers show how technological developments can result in conceptual ones, although Krohs and Callebaut are rather concerned about the results of these developments, while O'Malley and Soyer are more optimistic.<sup>4</sup>

---

<sup>4</sup> O'Malley and Soyer give examples of how technological developments, such as flow cytometry, have given rise to new biological topics, such as noise biology.



The papers have all provided us with new conceptual tools to help understand the changes we are witnessing in the life sciences; tools that will not only be useful for commentators on science such as philosophers, historians and sociologists, but also for scientists themselves. These conceptual tools have arisen out of a sensitivity to the changing nature of the life sciences, which means they are likely to be refined and modified as the science itself changes and develops.

Krohs explains how convenience experimentation, enabled by high-throughput data gathering methods, can lead to shifts in epistemic goals and in biological theorizing. O'Malley and Soyer's notion of integration elucidates current research practices, and, I have argued, could be expanded to encompass social and perhaps even policy dimensions of systems biology. Callebaut's scientific perspectivism shows how complex systems may require a broader range of perspectives than simpler ones, and points to the importance of distributed cognition in heterogeneous scientific communities. And I have suggested we add serious consideration of the nature of engineering to our theoretical repertoire.

What is most important in all three of these stimulating papers, however, is that "new ways of thinking about scientific practice are emerging" (O'Malley and Soyer p.30). It is clear that the categories of data-driven and hypothesis-driven research do not capture everything we want to explain when thinking about current biology, and that the concepts introduced in these papers will have an important role to play in future work.

## References

- Andrianantoandro E., Basu S., Karig D.K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, doi: 10.1038/msb4100073
- Brent, R. (2004). A partnership between biology and engineering. *Nature Biotechnology*, 22, 10, 1211-1214.
- Burian, R.M. (1997). Comments on Complexity and Experimentation in Biology. *Philosophy of Science*, 64, S279-S291.
- Callebaut, W. (this volume). Engineering life: changing the living world without trying to understand it. *Studies in History and Philosophy of Biological and Biomedical Sciences*
- Calvert, J. (2010). Systems biology, interdisciplinarity and disciplinary identity. In Parker, J.N., Vermeulen, N & Penders, B. (eds.) *Collaboration in the New Life Sciences* Aldershot: Ashgate.
- Calvert, J. and Fujimura, J. (2011). Calculating life: Duelling discourses in interdisciplinary systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 42, 2, 155-163.
- Endy, D. (2005). Foundations for engineering biology. *Nature*, 438, 449-453.
- Fisher, E., Mitcham, C., and Mahajan, R. (2006). Midstream modulation of technology: governance from within. *Bulletin of Science, Technology and Society* 26: 485–496.
- Giere, R.N. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press
- Heinemann M. and Panke, S. (2006). Synthetic biology – putting engineering into biology. *Bioinformatics*, 22, 22, 2790-2799.
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14, 3, 575-599.
- Keller, E.F. (2009). What Does Synthetic Biology Have to Do with Biology? *BioSocieties*, 4, 291-302.
- Royal Academy of Engineering (2009). *Synthetic Biology: Scope, Applications and Implications*. London: Royal Academy of Engineering.
- Krohs, U. (2010). Epistemic consequences of two different strategies for decomposing biological networks. In M. Suárez et al. (eds) *EPSA: Philosophical issues in the sciences*. Dordrecht: Springer.
- Krohs, U. (this volume). Convenience experimentation. *Studies in History and Philosophy of Biological and Biomedical Sciences*
- Mattila, E. (2005). Interdisciplinarity “in the making”: modeling infectious diseases. *Perspectives on Science*, 13, 4, 531-553.
- National Research Council (2009). *A New Biology for the 21st Century*, Washington DC: National Academies Press online at: <http://www.nap.edu/catalog/12764.html>
- O’Malley, M.A., Elliott, K.C., Haufe, C., and Burian, R.M. (2009). Philosophies of funding. *Cell*, 138, 611-615.
- O’Malley, M. and Soyer, O.S. (this volume). The dynamics of integration in molecular life science. *Studies in History and Philosophy of Biological and Biomedical Sciences*
- Steinle, F. (1997). Entering New Fields: Exploratory Uses of Experimentation. *Philosophy of Science*. 64, S65-S74.

- Van Fraassen, B.C. (2008). *Scientific representation: paradoxes of perspective*. Oxford: Oxford University Press.
- Vincenti, W.G. (1990). *What Engineers Know and How They Know It*. Johns Hopkins University Press.
- Wimsatt, W.C. (2007). *Re-engineering philosophy for limited beings: piecewise approximations to reality*. Cambridge, MA, Harvard University Press.
- Woese, C.R. (2004). A new biology for a new century. *Microbiology and Molecular Biology Reviews*, 68, 173-186.